

# Hengyu Zhao

University of California, San Diego  
Phone: +1-831-226-6427, Email: h6zhao@eng.ucsd.edu  
Office: EBU3B, Room 3256, 9500 Gilman Drive, La Jolla, CA 92093  
Personal website: <https://hengyuzhao.weebly.com/>

## EDUCATION

---

<b>University of California, San Diego, CA</b> Ph.D. in Computer Science and Engineering Advisor: Prof. Jishen Zhao	<i>03/2018 - 06/2021(expected)</i>
<b>University of California, Santa Cruz, CA</b> Ph.D. in Computer Engineering	<i>09/2016 - 03/2018</i>
<b>Xi'an Jiaotong University, Shaanxi, China</b> M.S. in Microelectronics	<i>09/2013 - 06/2016</i>
<b>Xi'an Jiaotong University, Shaanxi, China</b> B.S. in Microelectronics,	<i>09/2009 - 06/2013</i>

### Programming Skills

C++, Python, CUDA

## RESEARCH INTERESTS

---

- I am interested in the computer architecture and system design for autonomous driving systems and machine learning applications.

## RESEARCH AND WORK EXPERIENCE

---

### Work Experience

**Summer research intern at NVIDIA, Santa Clara, CA** *06/2020 - 09/2020*

This work analyzed the perception needs for AV safety. To study this, this work

- Built a framework to corrupt perception quality, by changing perception parameters (e.g., camera frame rate, neural network precision) and injecting errors to perception stack.
- Selected several common scenarios and tested how different perception qualities affect AV safety. Studied the correlation between scenario difficulty and safety.
- Proposed safety evaluation metrics to examine if the AV changed its trajectory, actuators or collided with obstacles.
- Proposed a safety sensitivity model that calculated the sensitivity of each perception parameter to AV safety.

**Summer research intern at Pony.ai, Fremont, CA** *06/2019 - 09/2019*

- Identified the memory traffic bottleneck of the ScanMatch module in the autonomous driving system, and designed a compute framework for such modules.
- Designed a multi-level tiling storing methodology that compressed the high definition map with different granularity.
- Based on the field of view (FOV) of LiDARs, proposed a FOV aware static map prefetch scheme to reduce the memory accesses. Verified this scheme with CUDA programming on GPUs, and designed a simulator to estimate the memory accesses.

**Summer research intern at Pony.ai, Fremont, CA** *07/2018 - 09/2018*

- Proposed a safety metric to guide the design of safety aware autonomous driving systems. The safety metric is a function of the perception latency of the system.
- Quantitatively analyzed the correlation between the perception latency and the obstacles around the autonomous vehicle, and built a latency model that helped the autonomous driving system to predict the latency in real-time.
- Exploited a case study, a resource management scheme, to demonstrate the effectiveness of the proposed safety metric and latency model. (*arXiv, IV'20, ICCD'20*)

## Research Experience

### Efficient computer architecture design for deep neural networks.

- Designed a heterogeneous processing in memory architecture to accelerate the neural network training. Proposed programming models and runtime systems to improve the programmability and hardware utilization. (*MICRO'18*)
- Leveraged the multi-level STT-RAM to perform the neural network training, and proposed a feature map coding scheme and a bit-wise write skipping scheme. This design reduced the memory capacity utilization and the energy consumption. (*MEMSYS'18*)

### Energy efficient hardware design for image and video processing.

- Leveraged the multi-level STT-RAM to efficiently process the images and videos, and improved the energy efficiency. (*ICCAD'17*)

## PUBLICATIONS

---

- [1] **Hengyu Zhao**, Yubo Zhang, Pingfan Meng, Hui Shi, Li Erran Li, Tiancheng Lou, Jishen Zhao. "Driving Scenario Perception-Aware Computing System Design in Autonomous Vehicles", *IEEE International Conference on Computer Design (ICCD)*, Oct, 2020. (**Best paper in track**)
- [2] **Hengyu Zhao**, Yubo Zhang, Pingfan Meng, Hui Shi, Li Erran Li, Tiancheng Lou, Jishen Zhao. "Safety Score: A Quantitative Approach to Guiding Safety-Aware Autonomous Vehicle Computing System Design", *IEEE Intelligent Vehicles Symposium*, Oct, 2020.
- [3] Sabur Baidya, Yu-Jen Ku, **Hengyu Zhao**, Jishen Zhao, Sujit Dey. "Vehicular and Edge Computing for Emerging Connected and Autonomous Vehicle Applications", *ACM/IEEE Design Automation Conference (DAC)*, July, 2020. (Invited Paper)
- [4] Saransh Gupta, Mohsen Imani, **Hengyu Zhao**, Fan Wu, Jishen Zhao, Tajana Rosing. "Implementing Binary Neural Networks in Memory with Approximate Accumulation", In *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, Aug. 2020.
- [5] **Hengyu Zhao**, Yubo Zhang, Pingfan Meng, Hui Shi, Li Erran Li, Tiancheng Lou, and Jishen Zhao. "Towards Safety-Aware Computing System Design in Autonomous Vehicles", In *arXiv repo*
- [6] **Hengyu Zhao**, Jiawen Liu, Matheus A. Ogleari, Dong Li, Jishen Zhao. "Processing-in-Memory for Energy-efficient Neural Network Training: A Heterogeneous Approach", *International Symposium on Microarchitecture (Micro)*, Oct. 2018. (Acceptance rate: 74/348=21%)
- [7] **Hengyu Zhao**, Jishen Zhao. "Leveraging MLC STT-RAM for Energy-efficient CNN Training", *Annual International Symposium on Memory Systems (MEMSYS)*, Oct. 2018.
- [8] **Hengyu Zhao**, Linuo Xue, Ping Chi, Jishen Zhao. "Approximate Image Storage with Multi-level Cell STT-MRAM Main Memory", *International Conference On Computer Aided Design (ICCAD)*, Nov. 2017. (Acceptance rate: 105/399=26%)
- [9] **Hengyu Zhao**, Colin Weinshenker, Mohamed Ibrahim, Adwait, Jog, Jishen Zhao. "Layer-wise Performance Bottleneck Analysis of Deep Neural Networks", *Workshop on Architectures for Intelligent Machines (AIM)*, Sep. 2017.
- [10] **Hengyu Zhao**, Hongbin Sun, Tai Min, Nanning Zheng. "Exploring the Use of Volatile STT-RAM for Energy Efficient Video Processing", *Proceedings of IEEE International Symposium on Quality Electronic Design (ISQED)*, March 2016. (**Best paper candidate**)
- [11] Qiubo Chen, **Hengyu Zhao**, Hongbin Sun, Nanning Zheng. "Exploiting bit-depth scaling for quality-scalable energy efficient display processing", *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2015.